

MODELADO ESTADÍSTICO DEL CAUDAL MENSUAL EN LA BAJA CUENCA DEL PLATAMelanie Meis^{1,2} y María Paula Llano^{1,2}¹Departamento de Ciencias de la Atmósfera y los Océanos, Universidad de Buenos Aires, Buenos Aires, Argentina²CONICET

(Manuscrito recibido el 21 de junio de 2017, en su versión final el 3 de noviembre de 2017)

RESUMEN

Los continuos estudios hidrológicos en la Cuenca del Plata son necesarios no solamente por ser la misma una fuente capaz de proveer energía a diferentes países en Sudamérica, navegabilidad y suministrar agua potable, sino también por los severos daños ocasionados por las frecuentes inundaciones que se producen en dicha cuenca. En este trabajo se propuso emplear una modelación estadística basada en modelos estacionales autoregresivos integrados de medias móviles (SARIMA), con el principal objetivo de representar y pronosticar las series temporales del caudal mensual en distintas estaciones de la baja Cuenca del Plata. Se plantearon modelos apropiados para la serie temporal del caudal mensual en la estación Timbúes (Río Paraná) para cinco subperíodos temporales de veinte años desde 1913 a 2012. Para las estaciones de Corrientes, Timbúes y Paso de los Libres se realizó un pronóstico a 32 meses (enero/2013-agosto/2015) teniendo en cuenta el último subperíodo: 1993-2012. La validación de los modelos fue incluida. Además, se realizó un estudio de distintos posibles escenarios a futuro en la estación Corrientes, contrastando los mismos con los valores registrados en ese período. En este sentido, con el análisis realizado, se pretendió encontrar modelos estadísticos que permitan contribuir al monitoreo del caudal mensual en la Cuenca del Plata, con el objetivo de asistir a los tomadores de decisión en la mitigación de los impactos negativos en la región.

Palabras clave: Cuenca del Plata; SARIMA; Pronóstico; Escenarios

STATISTICAL MODELING FOR THE MONTHLY STREAMFLOW IN THE SOUTH OF THE LA PLATA BASIN**ABSTRACT**

Hydrological studies in the La Plata basin are important as it is a source capable of providing water to several countries in South America, a large navigable network and rich in mineral resources. However, the necessity of these studies is highlighted by the frequent floods that occur in it. In this work, it is proposed to employ the statistical modelling through the Seasonal Autoregressive Integrated Moving Average (SARIMA) models, with the aim to represent and forecast the time series for monthly streamflow in different stations in the La Plata Basin. Proper models were proposed for the monthly streamflow time series in Timbúes station (Paraná

River) for five subtemporal periods of twenty years from 1913 to 2012. Moreover, for Corrientes, Timbúes and Paso de los Libres stations, a forecast for 32 months (January/2013-August/2015) was carried out taking into consideration the last subperiod, 1993-2012. The validation of the models was considered. Finally, forecast with different future scenarios were done in Corrientes station, and a comparison with the gauged values was included. In this way, it was pretended to find a statistical models in order to contribute to the monitoring of the monthly discharge in the La Plata Basin with the objective to mitigate negative impacts in the region.

Keywords: La Plata Basin; SARIMA; Forecast; Scenarios

1. INTRODUCCIÓN

El río Paraná es considerado uno de los ríos más importantes del mundo con una longitud aproximada de 3940 km., a su vez el río Uruguay presenta una extensión longitudinal menor y características fisiográficas distintas al anterior. Sin embargo, ambos ríos resultan de gran importancia en lo que refiere a la producción y la demanda hidroeléctrica, así como también en el suministro de agua a la actividad agrícola y ganadera en los países que integran la Cuenca del Plata. Por otra parte, estos ríos constituyen vías de comunicación y de transporte fluvial, es decir, actúan como conectores de centros de producción. Asimismo, dado el gran impacto que tienen ambos ríos en la sociedad, el monitoreo constante del río Paraná y el río Uruguay es necesario no solo para minimizar el riesgo de inundación, sino también para alertar sobre las bajantes que ocasionan inconvenientes en la toma de agua en diferentes localidades ribereñas. Es decir, la prevención, las redes de monitoreo, y los programas de alerta de eventos hidrológicos extremos implican a futuro importantes beneficios sociales, económicos y ambientales.

En las últimas décadas, la aplicación de modelos estadísticos en variables comprendidas en el área climatológica e hidrológica ha adquirido mayor demanda, no solo en el modelado de las mismas sino también para su pronóstico. Los modelos estocásticos como los autoregresivos de media móvil (ARMA), los modelos autoregresivos integrados de media

móvil (ARIMA) de Box y Jenkins (1976) o los modelos estacionales autoregresivos integrados de media móvil (SARIMA), son algunos de los utilizados debido a su bajo costo computacional como así también por su rápida implementación.

Diversos autores han considerado la metodología estadística desarrollada por Box y Jenkins en el estudio de distintas variables meteorológicas como temperatura, viento, precipitación y caudal, entre otras. Torres et al. (2005) realizaron un pronóstico del promedio horario de la velocidad del viento en Navarra, España. Concluyeron que estos modelos resultaban útiles en un pronóstico a corto plazo. Por otra parte, Cadenas y Rivera (2007) obtienen resultados adecuados en el pronóstico de la velocidad del viento en el sur de la costa de Oaxaca, México, para permitir luego una optimización de las plantas de generación de energía eléctrica.

Con respecto a la precipitación se aplicaron los modelos ARMA para obtener pronósticos a corto plazo en la cuenca del río Sieve en Italia (Toth et al., 2000) y en estaciones meteorológicas localizadas en el estado de Colorado (EE.UU) y el río Arno (Italia), (Burlando et al., 1992). En la India se desarrolló un modelado estadístico de precipitación mediante modelos ARIMA, Chattopadhyay y Chattopadhyay (2010) realizaron estudios con el Monzón de verano y Narayanan et al. (2013) con el período pre-Monzónico (marzo, abril y mayo).

En los últimos años se han implementado diferentes modelos estadísticos en el continente

africano. Estudios relacionados con los modelos SARIMA fueron realizados en distintas municipalidades de Ghana teniendo como objetivo obtener un modelo de precipitación mensual, como así también un pronóstico de la misma con el propósito de mitigar impactos negativos en el sector agricultor (Sampson et al., 2013). Afrifa-Yamohah et al. (2016) también consideraron los modelos SARIMA para el pronóstico mensual de la precipitación en Brong Ahafo, otra región de Ghana. Los autores encontraron un modelo que representaba el comportamiento de un modo significativo para un período de dos años, teniendo en cuenta su validación. Otro de los países en donde se aplicaron los modelos SARIMA fue en una región de Kenia (Kibunja et al., 2014), donde los autores seleccionaron un modelo apropiado de pronóstico de precipitación y su precisión para pronosticar la misma. En Warri Tower, Nigeria, el modelado de la precipitación y su pronóstico fue realizado con modelos estacionales ARIMA, encontrando un año de pronóstico adecuado (Eny y Adeyeye, 2015).

Por otro lado, el trabajo desarrollado por Mohammed (2015) podría ser un claro ejemplo de la aplicación de estos modelos estadísticos en el área hidrológica. En el mismo se pronosticó la cantidad de escorrentía para el año 2011 utilizando el promedio de todas las estaciones de medición dentro de cada estado de EE.UU, a partir de los modelos SARIMA y ARIMA. Por otro lado, en el mismo trabajo el autor sugiere que debido al cambio climático la escorrentía no puede ser pronosticada a más de un año.

Conocer el caudal de los ríos es de gran importancia para la navegación de los mismos, es por ello que De Figueiredo y Blanco (2016) modelaron el nivel del agua del río Amazonas utilizando modelos SARIMA con un horizonte de pronóstico a 24 meses encontrando resultados satisfactorios. Asimismo, en el río Caroní, Venezuela, el modelado de la serie temporal mensual del caudal en el periodo 1950-2003 fue realizado mediante modelos SARIMA (Bello et al., 2010), encontrando mediante un análisis de

los residuos un modelo adecuado.

En Argentina, los modelos autoregresivos (AR) fueron implementados por Pierini et al. (2012), en el pronóstico del caudal diario del río Colorado, los resultados obtenidos por esta metodología no resultaron tan precisos al momento de la validación.

El principal objetivo del presente trabajo fue seleccionar un modelo estocástico SARIMA (Wei, 2005) con el propósito de representar las series temporales del caudal mensual en tres estaciones de referencia: Timbúes, Corrientes y Paso de los Libres. Además, con una adecuada selección del modelo se pretendió desarrollar un pronóstico del caudal mensual. Finalmente, se plantearon distintos escenarios futuros, físicamente posibles, en la estación Corrientes, con el fin de definir un modelo que resulte una herramienta útil de monitoreo para los tomadores de decisión, permitiendo así la reducción de los impactos negativos en la baja Cuenca del Plata.

2. DATOS Y METODOLOGÍA

Los datos medios mensuales (período 1993-2015) de la Base de Datos Hidrológica Integrada (BDHI) publicada por la Subsecretaría de Recursos Hídricos de la República Argentina fueron considerados de las estaciones del río Paraná (Timbúes y Corrientes) y río Uruguay (Paso de los Libres) (Figura 1). Un primer estudio exploratorio fue realizado en la estación de Timbúes para el período 1913 -2012.

Desde la década del setenta un cambio en la tendencia del caudal fue establecido por García y Vargas (1998), y por Jaime y Menéndez (2002) en la Cuenca del Plata. Dada esta situación, se decidió analizar la presencia de este cambio en las series temporales utilizando el test de salto climático de Yamamoto (Yamamoto et al., 1987).

Por otra parte, el modelado estadístico empleado en este estudio consideró los modelos propuestos por Box and Jenkins (1976). Estos autores formularon modelos estocásticos para el estudio

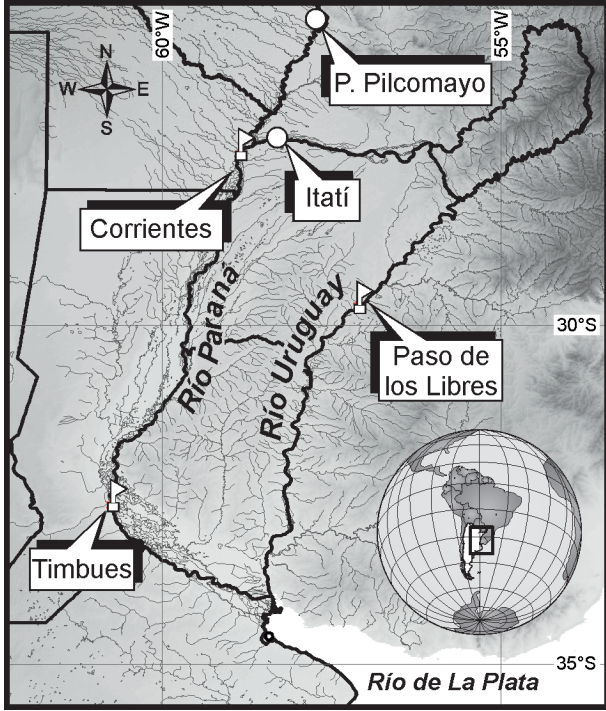


Figura 1: Ubicación geográfica de las estaciones de medición Corrientes, Timbúes, Paso de los Libres, Itatí y Puerto Pilcomayo en la Cuenca del Plata.

de las series de tiempo, como por ejemplo los modelos ARMA (Ec. 1)

$$\phi(B)Z_t = \theta(B)a_t + c \quad (1)$$

Donde $\phi(B)$ y $\theta(B)$ representan los términos asociados a la parte autoregresiva y de promedios móviles, respectivamente; c es un término constante en el modelo. Además, la Ec. (1) podría representarse de la siguiente manera:

$$\begin{aligned} Z_t - \phi_1 Z_{t-1} - \dots - \phi_p Z_{t-p} \\ = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \end{aligned} \quad (2)$$

con $a_t \sim WN(0, \sigma^2)$.

Los errores (a_t) siguen una distribución de ruido blanco con media cero y varianza σ^2 . Los parámetros p y q representan el orden de la parte autoregresiva (AR) y promedios móviles (MA), respectivamente.

Los ϕ_i , con $0 \leq i \leq p$, representan los parámetros estimados asociados a la parte AR, mientras que

los θ_i , con $0 \leq i \leq q$, están asociados a la parte MA. La combinación de ambos parámetros da lugar a los modelos ARMA.

Sin embargo, debido al supuesto de estacionaridad en las series temporales que estos modelos presentan, los modelos autoregresivos integrados de medias móviles (ARIMA) fueron desarrollados. Estos modelos permiten transformar las series temporales no estacionarias a estacionarias.

$$Z_t \sim ARIMA(p, d, q) \quad (3)$$

si $\phi_i(B)(1-B)^d Z_t = \theta_q(B)u_t$ y con d el orden de la diferencia a considerar ($d > 0$). Si $W_t = (1-B)^d Z_t$, el proceso $Z_t \sim ARIMA(p, q)$, con $u_t \sim WN(0, \sigma^2)$.

En particular, en este trabajo se implementó el modelado estadístico considerando la presencia de cierta estacionalidad en las series temporales utilizando los modelos estacionales autoregresivos integrados de medias móviles (SARIMA). Los modelos SARIMA permiten estudiar series temporales que no presentan un proceso estacionario, como así también series temporales que presentan cierta estacionalidad. Los modelos SARIMA se representan de la siguiente manera: SARIMA $(p, d, q)(P, D, Q)_S$, siendo (p, d, q) la parte ordinaria, y $(P, D, Q)_S$ la parte estacional de la serie de tiempo (implementado por Hyndman, 2016). En general, el modelo puede ser expresado de la siguiente manera (Ec. 4):

$$\begin{aligned} \phi_{is}(B_s)\phi_i(B)(1-B_s)^D(1-B)^d Z_t \\ = \theta_q(B_s)\theta_q(B)u_t \end{aligned} \quad (4)$$

con $\phi_i, 0 \leq i \leq p, \phi_{is}, 0 \leq i \leq P, \theta_i, 0 \leq i \leq q, \theta_{is}, 0 \leq i \leq P$.

$D > 0$ el orden de la diferencia asociada a la parte estacional del modelo.

Es importante mencionar, que en todos los casos se consideró p, q, P y Q menores o iguales a 2 para obtener así modelos parsimoniosos (modelos sencillos). En este sentido, considerar

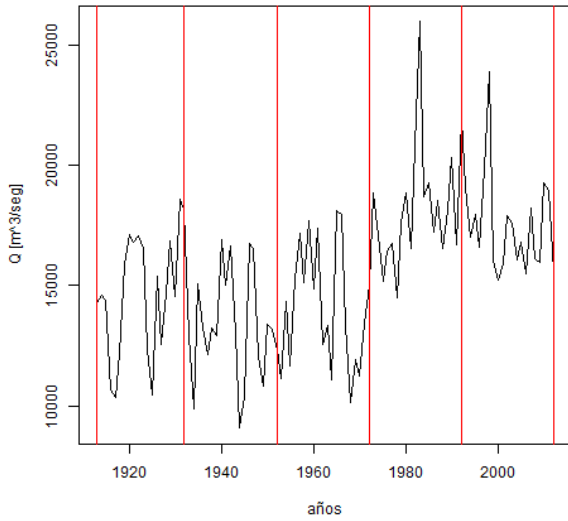


Figura 2: Serie temporal del caudal medio anual en la estación Timbúes para el período 1913-2012. En líneas rojas se marca los subperíodos: 1913-1932; 1933-1952; 1953-1972; 1973-1992; 1993-2012.

un modelo con un orden mayor de parámetros podría significar un estudio más complejo y más costoso computacionalmente.

Con el objetivo de modelar las series de caudal mediante el mejor modelo estadístico posible se consideraron métodos de identificación (función de autocorrelación (ACF), autocorrelación parcial (PACF)), estimación de los parámetros del modelo y métodos de diagnóstico. Para ello, el estudio de los residuos fue realizado mediante diferentes técnicas (ACF, PACF) y test estadísticos, como el de Ljung-Box con la hipótesis nula que los datos considerados se distribuyen de forma independiente (5 % de nivel de significancia). Para la selección del modelo, un cierto compromiso entre la ACF, PACF y el test de Ljung-Box fue establecido.

Además, si dos o más modelos resultaban ser representativos de las series temporales, se seleccionó al mejor mediante el criterio de Akaike (AIC), es decir, el modelo con menor valor de AIC. Como así también se tuvieron en cuenta otros

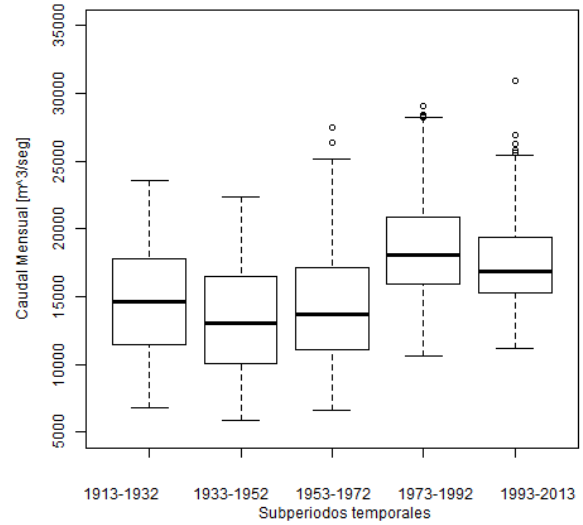


Figura 3: Boxplot correspondiente a cada subperíodo temporal estudiado para el caudal mensual en la estación Timbúes. Se muestran los principales cuartiles (25 %, 50 %, 75 %) junto con los valores extremos (círculos).

índices: el error medio (ME), error cuadrático medio (RMSE), error medio absoluto (MAE), error porcentual medio (MPE) y error absoluto porcentual medio (MAPE).

Finalmente, un pronóstico para el caudal mensual fue desarrollado para el período 01/2013 a 08/2015 (32 meses) considerando el subperíodo de 20 años (1993-2012) en Timbúes, Corrientes y Paso de los Libres (Ec. 5).

$$\begin{aligned} \phi_{is}(B_s)\phi_i(B)(1-B_s)^D(1-B)^d Z_{t+1} \\ = \theta_q(B_s)\theta_q(B)u_{t+1} \end{aligned} \quad (5)$$

Los datos desde enero 2013 a agosto 2015 fueron tenidos en cuenta para validar el pronóstico mensual del caudal (Stoffer, 2016).

3. RESULTADOS Y DISCUSIÓN

3.1. Análisis exploratorio en la estación Timbúes (río Paraná)

De acuerdo a estudios previos relacionados con

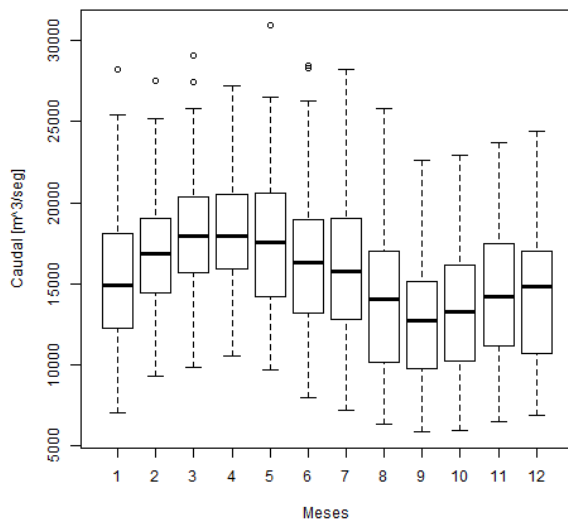


Figura 4: Boxplot correspondiente al caudal medio mensual para el período 1913-2012. Se muestran los cuartiles (25 %, 50 %, 75 %) junto con sus valores outliers (círculos).

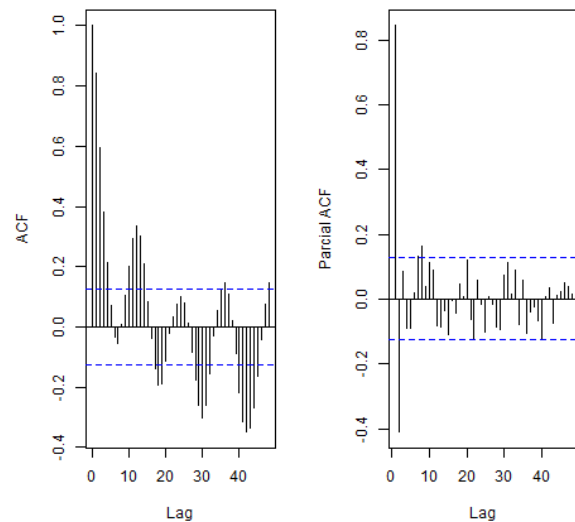


Figura 5: ACF (izq.) representa la función de autocorrelación y PACF (der.) representa la función de autocorrelación parcial de la serie de caudal mensual en la estación Timbúes para el subperíodo 1913-1932.

la tasa de precipitación, y en consecuencia a los cambios presentes en los valores del caudal en la Cuenca del Plata en las últimas décadas del siglo XX se analizó la presencia de un cambio en los valores medios de caudal anual de la estación Timbúes (Figura 2) mediante el test de Yamamoto. Los resultados obtenidos concuerdan con la bibliografía ya mencionada, ubicando un cambio notorio en el año 1973. De acuerdo a este resultado, la serie temporal fue dividida en cinco subperíodos de veinte años: 1913-1932; 1933-1952; 1953-1972; 1973-1992; 1993-2012.

Para sintetizar el comportamiento del caudal en cada período se realizaron gráficos de boxplot, en donde se muestran los percentiles: 25, 50 y 75, junto con el valor máximo, mínimo y valores outliers (Figura 3). A partir de la misma se observó un cambio en el valor medio, siendo más notorio desde el cuarto período. Además, se halló un mayor número de casos extremos para los últimos dos períodos temporales.

Por otro lado, se exploró la estacionalidad en

la serie de tiempo correspondiente a la estación Timbúes. En este sentido, fue clara la presencia de una onda anual (Figura 4), con el máximo medio mensual de caudal presente en los meses de marzo y abril, y el mínimo alrededor del mes de septiembre.

Dentro del estudio exploratorio del caudal mensual de la estación Timbúes para cada subperíodo se analizaron las autocorrelaciones y las autocorrelaciones parciales (se presentan solo los gráficos correspondientes al primer subperíodo). En la Figura 5 se exhiben los valores de la autocorrelación de la serie de caudal, los mismos resultaron significativos en la mayoría de los primeros 12 meses, lo que estaría implicando la presencia de una componente no estacionaria en la serie. Por tal motivo se decidió diferenciar la serie una vez y se calculó la función de autocorrelación de la serie diferenciada. En la Figura 6 es posible ver con mayor detalle la estacionalidad de la serie (onda anual), por este motivo la serie fue diferenciada estacionalmente.

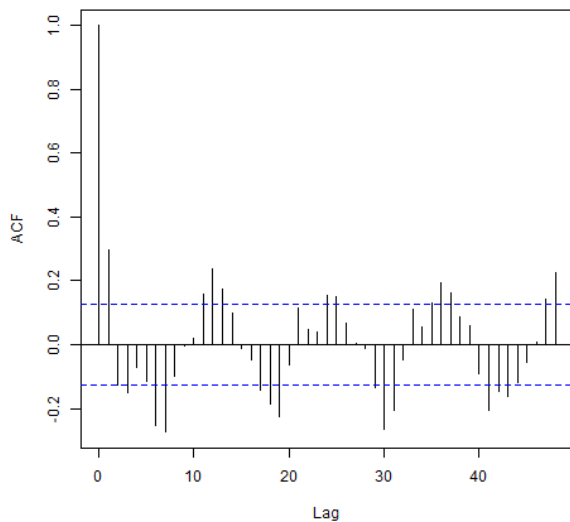


Figura 6: Función de autocorrelación (ACF) de la serie diferenciada en la estación Timbúes del caudal mensual correspondiente al período 1913-1932.

Una vez decidida la diferenciación estacionaria y estacional se plantearon todas las posibles combinaciones de p y q entre 0 y 2, y P y Q entre 0 y 1. Los 36 modelos obtenidos fueron estudiados considerando el análisis de los residuos con la función de autocorrelación y la función de autocorrelación parcial, y el test de Ljung-Box para establecer así los modelos más adecuados para cada subperíodo temporal. Para decidir el mejor modelo se calcularon los siguientes índices: AIC, ME, RMSE, MAE, MPE y MAPE. En la Tabla I se presentan los valores de estos índices para todos los modelos del primer subperíodo.

De acuerdo al análisis anterior, el modelo seleccionado para el primer subperíodo resultó ser el SARIMA $(1, 1, 2)(0, 1, 1)_{12}$. Igual procedimiento se realizó con los otros subperíodos, los modelos seleccionados fueron para el segundo período (1933-1952) el SARIMA $(2, 1, 2)(1, 1, 1)_{12}$ y para el tercer período (1953-1972) el modelo SARIMA $(1, 1, 2)(0, 1, 1)_{12}$.

El cuarto y el quinto período no pudieron ser

bien modelados por ninguna de las 36 posibles combinaciones consideradas en los subperíodos precedentes. Estos dos últimos períodos exhibieron la particularidad de presentar valores muy extremos (asociados a eventos Niño). Como una primera solución para poder modelar a estos subperíodos se decidió reemplazar los valores extremos por sus respectivos percentiles 95, para de esta forma suavizar el valor extremo sin ser eliminado del análisis y sin cambiar a una metodología robusta.

Con estas nuevas series se procedió al mismo análisis anterior, dando como resultado la particularidad de que en ambos subperíodos no fue necesario diferenciar la componente no estacional de la serie. Es decir, el modelo SARIMA $(2, 0, 2)(1, 1, 1)_{12}$ pudo ser considerado para modelar el cuarto período, con la salvedad de que se observó en los residuos un lag significativo a los 18 meses. Por otra parte, para el último subperíodo el modelo SARIMA $(1, 0, 2)(0, 1, 1)_{12}$ fue utilizado para modelar la serie temporal. En este caso, el lag 38 resultó significativo en el análisis de la ACF y la PACF de los residuos.

De acuerdo a los modelos previos seleccionados, en la Tabla II se presentan los coeficientes asociados a cada modelo para cada subperíodo.

3.2. Pronóstico SARIMA para la estación Timbúes

Habiendo mostrado que los cinco subperíodos temporales de la estación Timbúes no pudieron ser representados por un único modelo, se utiliza el quinto período (1993-2012) para el pronóstico de los siguientes 32 meses (01/2013 a 08/2015). El mismo procedimiento fue realizado en las estaciones Corrientes y Paso de los Libres.

El modelo SARIMA $(1, 0, 2)(0, 1, 1)_{12}$ se utilizó para el pronóstico mensual a 32 meses para la estación Timbúes (Figura 7). Además, se exhiben las bandas de un desvío estándar para representar un cierto intervalo de confianza y los valores reales medidos para mostrar la validación

p	d	q	P	D	Q	AIC	ME	RMSE	MAE	MPE	MAPE
0	1	0	0	1	0	4240,52	45,35	2668,20	2072,47	-0,73	16,12
1	1	0	0	1	0	4236,36	39,35	2632,10	2047,27	-0,57	15,93
2	1	0	0	1	0	4217,56	44,26	2513,21	1974,26	-0,59	15,16
0	1	1	0	1	0	4230,87	36,64	2600,16	2035,17	-0,47	15,80
1	1	1	0	1	0	4228,97	39,31	2577,75	2017,27	-0,51	15,58
2	1	1	0	1	0	4217,74	46,18	2503,01	1971,50	-0,62	15,17
0	1	2	0	1	0	4222,41	46,15	2540,45	1986,34	-0,64	15,23
1	1	2	0	1	0	4204,96	63,47	2414,99	1879,33	-1,04	14,60
2	1	2	0	1	0	4218,35	45,31	2494,93	1957,70	-0,61	15,06
0	1	0	1	1	0	4197,11	27,32	2400,43	1889,38	-0,89	14,45
1	1	0	1	1	0	4189,88	21,95	2350,88	1827,15	-0,67	13,95
2	1	0	1	1	0	4169,27	23,25	2235,14	1774,01	-0,75	13,49
0	1	1	1	1	0	4182,32	19,72	2311,13	1797,89	-0,57	13,68
1	1	1	1	1	0	4182,02	21,39	2299,90	1792,09	-0,63	13,61
2	1	1	1	1	0	4168,79	24,49	2222,45	1776,95	-0,81	13,52
0	1	2	1	1	0	4177,12	25,57	2275,61	1786,05	-0,79	13,53
1	1	2	1	1	0	4159,91	29,03	2161,57	1699,05	-1,39	13,12
2	1	2	1	1	0	4170,43	24,15	2220,64	1764,21	-0,77	13,45
0	1	0	0	1	1	4120,01	50,80	1882,33	1500,84	-0,84	11,47
1	1	0	0	1	1	4115,58	42,60	1855,74	1477,96	-0,69	11,29
2	1	0	0	1	1	4096,43	49,61	1770,59	1405,08	-0,71	10,73
0	1	1	0	1	1	4110,27	39,16	1833,94	1452,73	-0,61	11,10
1	1	1	0	1	1	4110,02	41,76	1824,87	1449,29	-0,65	11,05
2	1	1	0	1	1	4095,18	57,27	1757,74	1387,66	-0,68	10,64
0	1	2	0	1	1	4103,81	53,50	1799,84	1438,92	-0,77	10,98
1	1	2	0	1	1	4093,54	55,91	1746,21	1371,52	-1,03	10,57
2	1	2	0	1	1	4096,42	51,67	1754,46	1387,55	-0,68	10,65
0	1	0	1	1	1	4122,01	50,75	1882,19	1500,74	-0,84	11,47
1	1	0	1	1	1	4117,58	42,47	1855,32	1477,67	-0,69	11,29
2	1	0	1	1	1	4098,42	49,34	1769,72	1404,39	-0,71	10,73
0	1	1	1	1	1	4112,27	39,02	1833,46	1452,56	-0,61	11,09
1	1	1	1	1	1	4112,02	41,76	1824,88	1449,30	-0,65	11,05
2	1	1	1	1	1	4097,12	56,78	1756,07	1386,99	-0,69	10,64
0	1	2	1	1	1	4105,81	53,60	1800,06	1439,08	-0,77	10,98
1	1	2	1	1	1	4095,54	56,10	1746,43	1371,74	-1,03	10,57
2	1	2	1	1	1	4098,38	51,20	1753,04	1386,75	-0,69	10,65

Tabla I: Criterio de Akaike (AIC), el error medio (ME), error cuadrático medio (RMSE), error medio absoluto (MAE), error porcentual medio (MPE) y error absoluto porcentual medio (MAPE) para las 36 posibles combinaciones de los modelos SARIMA del caudal medio mensual en el período 1913-1932 para la estación Timbúes. En negrita se destaca el modelo seleccionado SARIMA (1, 1, 2)(0, 1, 1)₁₂.

Período		1913-1932	1933-1952	1953-1972	1973-1992	1993-2012
Modelo SARIMA		(1, 1, 2) (0, 1, 1)	(2, 1, 2) (1, 1, 1)	(1, 1, 2) (0, 1, 1)	(2, 0, 2) (1, 1, 1)	(1, 0, 2) (0, 1, 1)
Coeficientes	ar1	0,676 (0,111)	0,502 (0,224)	0,649 (0,086)	1,414 (0,350)	0,636 (0,111)
	ar2		0,240 (0,204)		-0,474 (0,272)	
	ma1	-0,540 (0,111)	-0,421 (0,202)	-0,562 (0,094)	-0,467 (0,341)	0,476 (0,141)
	ma2	-0,361 (0,063)	-0,580 (0,201)	-0,362 (0,070)	-0,210 (0,089)	0,102 (0,092)
	sar1		0,161 (0,086)		0,103 (0,087)	
	sma1	-1,000 (0,057)	-0,850 (0,064)	-1,000 (0,076)	-0,900 (0,073)	-0,904 (0,065)

Tabla II: Coeficientes de los parámetros de los modelos SARIMA seleccionados para cada uno de los subperíodos en la estación Timbúes. (Entre paréntesis su respectivo desvío estándar).

del modelo. Por lo observado, el modelo podría considerarse preciso hasta el año y medio. Sin embargo, los meses de medición correspondientes a julio y agosto no están dentro del intervalo de confianza. En relación a esto, en la Figura 8 se presenta la distribución del máximo absoluto mensual. A partir de la misma se observa que la mayor frecuencia de ocurrencia resultó ser durante la estación de otoño (marzo y abril). Pero, en un 10 % de los años el valor máximo mensual sucedió durante el mes de julio. Así mismo, según el informe del Instituto Nacional del Agua (Goniadzki et al., 2013a) en el mes de julio de 2013 se produjo la entrada inesperada de una onda de crecida que llevo al aumento del caudal y en consecuencia al aumento en la altura del río, dando lugar a alertas de riesgo de inundación y evacuaciones en varios sectores de la Cuenca del Plata. Es necesario mencionar la situación registrada en el mes de febrero de 2013, donde el modelo representó un continuo aumento en el caudal durante los primeros meses. Mientras que la situación registrada fue una bajante en el caudal debido a las anomalías de precipitación ocurridas en los meses precedentes (Goniadzki et al., 2013b).

3.3. Pronóstico SARIMA para las estaciones de Corrientes y Paso de los Libres

Para la estación Corrientes (período 1993-2012) el modelo SARIMA (1, 1, 2)(1, 1, 1)₁₂ fue

seleccionado. Para este modelo, una previa atenuación de los valores outliers debió ser realizada, y la diferencia ordinaria fue considerada ($d=1$). Sin embargo, en el análisis de los residuos de la ACF y la PACF, los lags 15 y 26 presentaron cierta significancia (Figura 9 y Figura 10). Por otra parte, los p-valores correspondientes al test de Ljung-Box resultaron ser mayores a 0,05 (Figura 11). El pronóstico a 32 meses fue realizado teniendo en cuenta el modelo anterior (Figura 12), observándose que en los meses de febrero, julio y agosto el pronóstico no representó fehacientemente a los valores medidos, en concordancia a la bajante y crecida mencionada en el ítem anterior.

Para la estación de Paso de los Libres (río Uruguay) el modelo SARIMA (1, 0, 2)(0, 1, 1)₁₂ fue encontrado como el modelo que mejor representó esta serie temporal. La única diferencia con el análisis en la estación Corrientes fue que en este caso, la atenuación de los valores extremos no presentó ninguna mejoría en el modelado, por tal motivo no fue tenida en cuenta. De acuerdo al análisis de la ACF de los residuos, no se encontraron lag significativos para los primeros 40, sin embargo, en el estudio de la PACF el lag 10 resultó significativo. De todas maneras, los p-valores asociados al test de Ljung-Box superaron el valor 0,05 para los primeros 40 lag. Luego, con la selección del modelo, el pronóstico del caudal mensual a 32

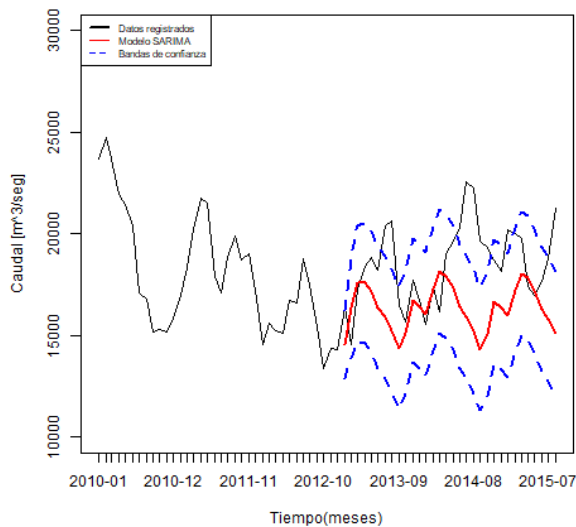


Figura 7: Pronóstico y validación para la estación Timbúes mediante el modelo SARIMA $(1, 0, 1)(0, 1, 1)_{12}$ (línea roja). Líneas punteadas azules representan un desvío respecto del valor pronosticado. En líneas negras los valores observados de caudal medio mensual.

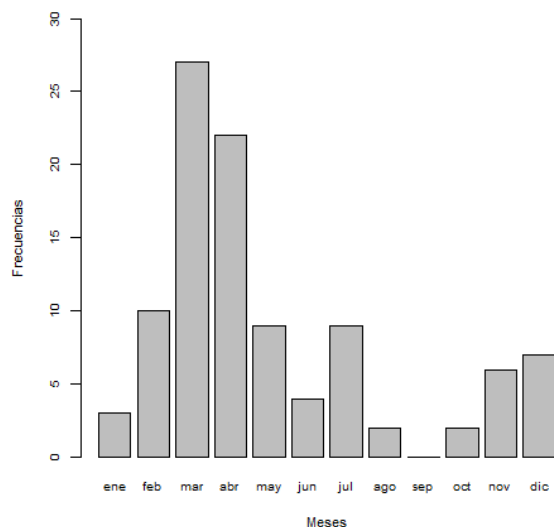


Figura 8: Distribución de la ocurrencia del máximo absoluto de caudal mensual en Timbúes para el período 1913-2012.

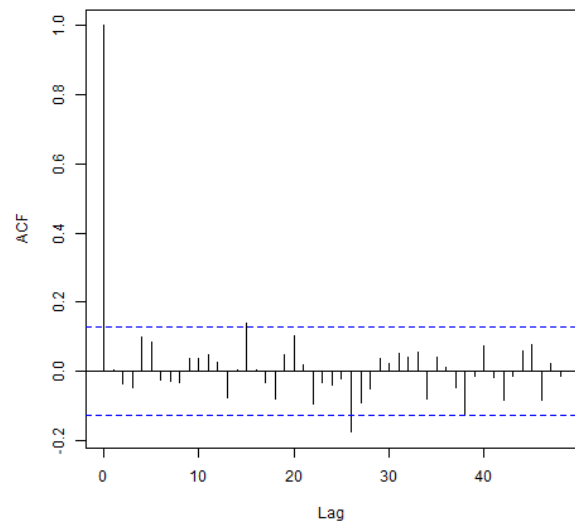


Figura 9: Función de autocorrelación (ACF) de los residuos del modelo SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ del caudal mensual correspondiente al período 1993-2012 en la estación Corrientes.

meses en esta estación fue realizado (Figura 13). Se observó que tanto enero como abril, no están dentro del intervalo de confianza (valor modelado a futuro \pm un desvío estándar). Sin embargo, teniendo en cuenta dos desvíos estándar dichos valores entrarían dentro del intervalo. Sin tener en cuenta estos dos meses, el pronóstico sería prácticamente preciso para al menos un año y medio (plazo de tiempo concordante con la bibliografía existente).

3.4. Modelado exploratorio de distintos escenarios en la estación Corrientes

En base a los resultados encontrados para el modelado de las series se decidió plantear para la estación Corrientes, en el río Paraná, distintos escenarios para estos 32 meses incorporando variables regresoras al modelo considerado previamente (ítem 3.3). Dado que la estación Itatí está situada aproximadamente a 61 km. de la estación Corrientes, se plantearon distintas magnitudes del caudal mensual en dicha estación. Se seleccionaron los años en donde el caudal

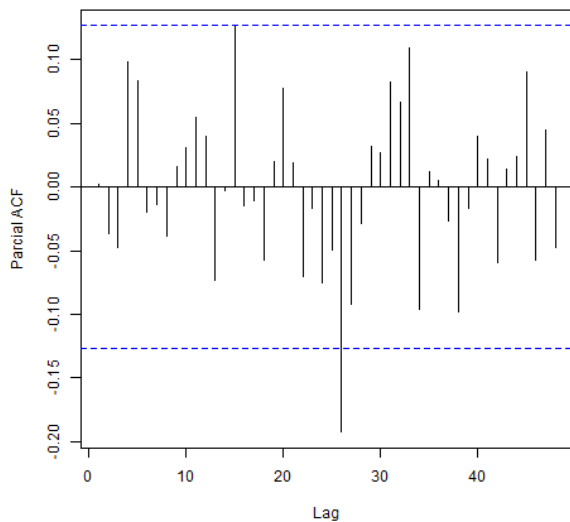


Figura 10: Función de autocorrelación parcial (PACF) de los residuos del modelo SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ del caudal mensual correspondiente al período 1993-2012 en la estación Corrientes.

anual acumulado del período 1993-2012 fuese máximo y mínimo. Además, se incluyó en el estudio la onda media de caudal. Se consideró también el caudal de uno de los ríos afluentes (Paraguay) medido en la estación Puerto Pilcomayo, seleccionando nuevamente los años de máximo y mínimo, y la onda media de caudal. Una vez identificados los años se tomaron los valores mensuales correspondientes. Se simularon las 3 situaciones de caudal máximo (Q_{max}), mínimo (Q_{min}) y medio (Q_{med}). Como así también se planteó el escenario de persistencia con el caudal de los últimos 12 meses (Q_{12}) en ambas estaciones.

La Figura 14 muestra el caudal simulado para los diferentes escenarios, en conjunto con los valores registrados en el período 1/2013- 8/2015 en la estación Corrientes. Por un lado se observa, que al considerar el modelo inicial SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ sin incluir variables regresoras o el escenario Q_{med} , ambos no presentaron diferencias. Lo anterior es de esperar ya que el modelo inicial está representando

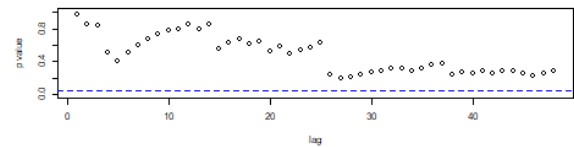


Figura 11: P-values del test Ljung-Box de los residuos del modelo SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ en la estación Corrientes para el período 1993-2012.

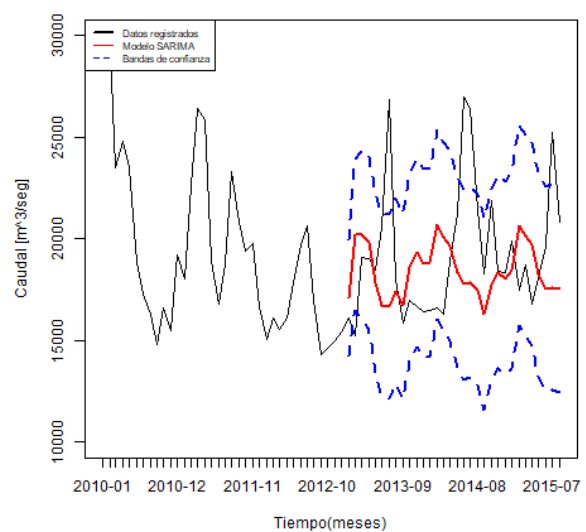


Figura 12: El caudal mensual pronosticado para la estación Corrientes mediante el modelo SARIMA $(1, 1, 2)(1, 1, 1)_{12}$ (línea roja). Líneas punteadas azules representan un desvío estándar respecto del valor pronosticado. En líneas negras los valores observados de caudal medio mensual.

el comportamiento medio del caudal en la estación Corrientes lo que estaría incluyendo el comportamiento medio del caudal en Pilcomayo e Itatí.

Al considerar el escenario Q_{12} , el pronóstico presentó una notable mejoría con respecto al modelo inicial. Es posible apreciar que el modelo logra simular la ocurrencia del máximo y del mínimo, con algunas subestimaciones dada las características extraordinarias de la crecida

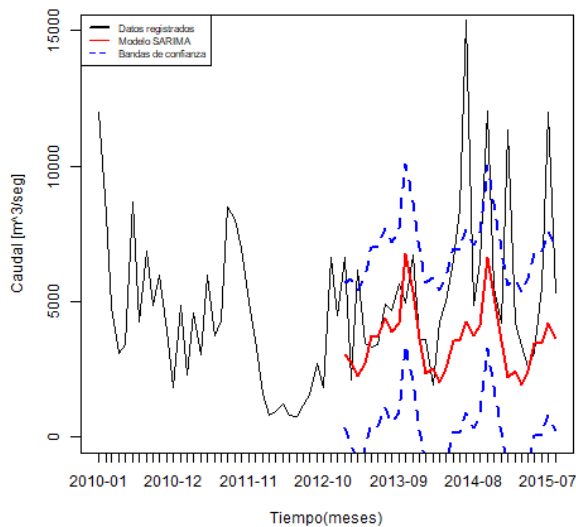


Figura 13: El caudal mensual pronosticado para la estación Paso de los Libres mediante el modelo SARIMA $(1, 0, 2)(0, 1, 1)_{12}$ (línea roja). Líneas punteadas azules representan un desvío estándar respecto del valor pronosticado. En líneas negras los valores observados de caudal medio mensual.

ocurrida en julio/agosto de 2013 (ya mencionada en el ítem 3.2).

En el escenario de los mínimos caudales (Q_{min}) se modeló con un mínimo error lo realmente ocurrido en los primeros tres meses del año. Esto se debió a que los registros del río Paraguay en la estación Puerto Pilcomayo y en el río Paraná en la estación Itatí presentaron valores cercanos al escenario considerado de valores mínimos extremos.

Es interesante destacar que estos tres escenarios (Q_{med} , Q_{min} , Q_{12}) se encuentran en un rango medio de valores de 4000 m³/seg. Dicha variabilidad se encuentra dentro de la variabilidad intercuartil propia de la serie de caudal mensual en la estación Corrientes.

Por otra parte, en la simulación del escenario extremo de caudal máximo Q_{max} , se obtuvieron

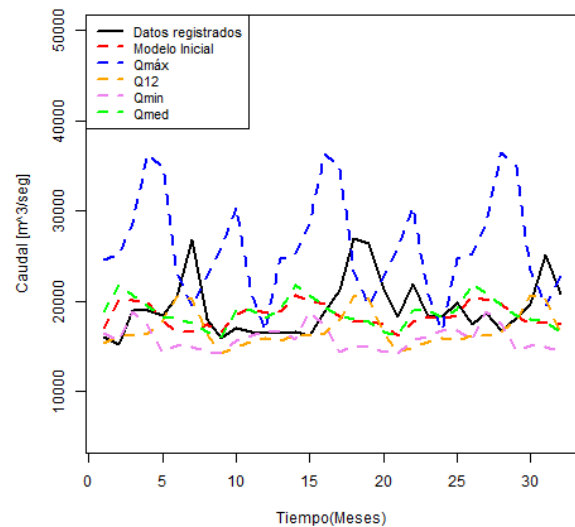


Figura 14: Diferentes escenarios de pronóstico en la estación Corrientes para el período 1/2013-8/2015. (Modelo inicial, máximo caudal ($Q_{máx}$), mínimo caudal ($Q_{mín}$), caudal medio (Q_{med}) y caudal persistente (Q_{12})).

valores superiores a los 35000 m³/seg en los meses de abril y mayo. Este escenario estaría representando el mayor caudal registrado en los últimos 20 años, debido a que las variables regresoras emplean los datos del año 1998. En dicho año en la estación Corrientes se registró un caudal medio en abril superior a los 40000 m³/seg.

4. CONCLUSIONES Y DISCUSIÓN

El continuo análisis hidrológico en la Cuenca del Plata es necesario dado el potencial riesgo de valores extremos de caudal (inundaciones y bajantes) que provocan daños irreparables en regiones dentro de la cuenca. En este sentido, el monitoreo del caudal mensual es fundamental. Para lo cual, el modelado estadístico resultaría una herramienta útil para mejorar el sistema de alerta y control en el futuro cercano. De esta manera, en este trabajo se empleó el modelado estadístico de acuerdo a los modelos SARIMA, propuesto por Box y Jenkins, para

ciertas estaciones localizadas en la parte baja de la Cuenca del Plata.

El salto en los valores de caudal anual a principios de la década del 70 fue corroborado en la estación Timbúes, en base al mismo y a la evolución temporal de la serie, el estudio del caudal mensual se decidió realizarlo considerando subperíodos temporales de veinte años.

A través de herramientas de identificación, estimación y diagnóstico se propuso modelar las series de 20 años de la estación Timbúes. Para ello, todas las posibles combinaciones de modelos SARIMA con p y q entre cero y dos, P y Q entre cero y uno fueron contempladas y se seleccionó mediante diferentes índices el mejor modelo para cada subperíodo. Por otra parte fue posible detectar la influencia que implica incluir valores outliers al modelado. Con respecto a los valores extremos, el reemplazo de los mismos por su correspondiente percentil 95 de cada subperíodo temporal permitió una mejor selección del modelo SARIMA en las estaciones del río Paraná.

Por otro lado, es importante mencionar que un único modelo para los cinco subperíodos temporales en la estación Timbúes no fue posible, probablemente relacionado con los cambios que ha sufrido el caudal durante los últimos 100 años. Muchos de esos cambios relacionados con el uso del suelo, el cambio climático, y también con la construcción de presas (Tucci et. al, 2009; Saurral et al., 2008).

El pronóstico del caudal mensual realizado para las estaciones de Timbúes, Corrientes y Paso de los Libres para 32 meses presentó ciertas singularidades. Parte de estas singularidades podrían estar relacionadas con ondas de crecida inesperadas para la época del año. De todas maneras, este pronóstico resulta prácticamente óptimo hasta el año y medio.

Por otra parte, el planteo de los diferentes modelos SARIMA en la estación Corrientes permitió reconocer que el modelo autoregresivo simple presentó ciertas inconsistencias

subestimando y sobreestimando los valores reales. Al incorporar variables regresoras, es decir, teniendo en cuenta valores de caudal mensual en dos estaciones predecesoras a la estación Corrientes se plantearon distintos escenarios de posibles estados del caudal. El escenario de caudal medio no exhibió notables diferencias respecto al modelo simple. El caudal máximo (Q_{max}) es un escenario extremo que sobrestima a los valores observados, sin embargo, dado que los valores considerados para las variables regresoras representaron valores reales de medición, dicho escenario sería útil de estudiar, si se pronostica un fenómeno ENSO de igual o mayor magnitud que el ocurrido en el año 1998. Por otra parte, el escenario de caudal mínimo (Q_{min}) logra estimar los primeros meses del pronóstico pero luego subestima a los valores reales. El modelo persistente (Q_{12}) es el que mejor representó el máximo desplazado, que se presentó en el mes de julio. Este escenario permite reconocer un retraso en la ocurrencia del máximo mensual hacia los meses de junio y julio.

En base a los análisis realizados, el estudio de modelos estadísticos particularmente los modelos SARIMA incluyendo variables regresoras podría resultar una herramienta útil para modelar las series temporales, más aún para el pronóstico de las mismas. Estos modelos debido a su rápida implementación, podrían contribuir a la reducción de impactos negativos causados por importantes inundaciones en regiones de intereses económicos y sociales cercanos a los ríos que se proponen en este estudio.

Es importante aclarar que los modelos estadísticos no son las únicas metodologías para prever una mitigación de impactos negativos. Los modelos hidrodinámicos constituyen una parte fundamental para complementar el análisis anterior no solo debido a la incorporación de ecuaciones físicas sino también a la inclusión de la topografía (Jaime y Menéndez, 1997 y Re et al., 2015).

Agradecimientos: A la Subsecretaría de Recursos Hídricos. A los proyectos de investigación PIP 112-201-301-00806 del CONICET, 2002130300014BA de la UBA. A los aportes de la Dra. Elena Martinez. A los dos revisores anónimos por sus valiosos comentarios y discusiones sobre el manuscrito.

REFERENCIAS

- Afrifa-Yamohah E., Bashiru I., Saeed, I., Azumah, K., 2016. Sarima Modelling and Forecasting of Monthly Rainfall in the BrongAhafo Region of Ghana. *WorldEnvironment* 6(1), 1-9.
- Bello, J., Vega González, C., Perez, E., 2010. Análisis estocástico de una serie cronológica de caudales mensuales. Caso de estudio: Río Caroní, Venezuela. *Revista Ingeniería UC*, Vol. 17, No. 1, 46-54.
- Box, G., Jenkins, G., 1976. *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco, CA.
- Burlando, P., Rosso, R., Cadavid, L., Salas, J., 1992. Forecasting of short-term rainfall using ARMA models. *Journal of Hydrology* 144, 193-211.
- Chattopadhyay, S., Chattopadhyay, G., 2010. Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN. *C. R. Geoscience* 342, 100-107.
- Cadenas, E., Rivera, W., 2007. Wind speed forecasting in the SouthCoast of Oaxaca, México, *Renewable Energy*. Vol. 32, no. 12, pp 2116-2128.
- De Figueiredo, N., Cavalcante Blanco, C., 2016. Water level forecasting and navigability conditions of the Tapajós River- Amazon-Brazil. *La Houille Blanche*. Vol. 3, 53-64.
- Eny, D., Adeyeye, F., 2015. Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection*. Vol. 3, 91-98.
- García, NO., Vargas, WM., 1998. The temporal climatic variability in the "Rio de la Plata" basin displayed by the rivers discharges. *ClimaticChange*. Vol. 38, 359-379.
- Goniadzki, D., Borús, J., Almeida, G., Díaz, L., Núñez, V., Contreras, G., 2013a. Posibles escenarios hidrológicos en la Cuenca del Plata durante el período agosto-septiembre-octubre de 2013. Instituto Nacional del Agua (INA).
- Goniadzki, D., Borús, J., Almeida, G., Díaz, L., Núñez, V., Contreras, G., 2013b. Posibles escenarios hidrológicos en la Cuenca del Plata durante el período marzo-abril-mayo 2013. Instituto Nacional del Agua (INA).
- Hyndman, R., 2016. Forecasting Functions for Time Series and Linear Models. Package 'forecast'. <http://github.com/robjhyndman/forecast>.
- Jaime, P., Menéndez, A., 1997. Modelo hidrodinámico del río Paraná desde Yaciretá hasta la Ciudad de Paraná. Informe. Instituto Nacional del Agua (INA).
- Jaime, P., Menéndez, A., 2002. Análisis del Régimen Hidrológico de los ríos Paraná y Uruguay. Informe INA-LHA 05-216-02, Comitente: Proyecto Freplata.
- Kibunja, H., Kihoro, J., Orwa, G., Yodah, W., 2014. Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region. *Mathematical Theory and Modeling*. Vol. 4, No. 11.
- Mohammed, V., 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorol. Appl.* 22, 592-598.
- Narayanan, P., Basistha, A., Sarkar, S., Sachdeva, K., 2013. Trend analysis and ARIMA modelling of pre-monsoon rainfall data for Western India. *C. R. Geoscience* 345, 22-27.
- Pierini, J., Gómez, E., Telesca, L., 2012. Prediction of water flows in Colorado River, Argentina. *Lat. Am. J. Aquat. Res.* 40(4), 872-880.
- Re, M., Sabarots Gerbec, M., Storto, L., 2015. Estadística de niveles en el Delta del río Paraná mediante modelación hidrodinámica. Séptimo simposio sobre regional sobre hidráulica de ríos.
- Sampson, W., Suleman, N., Gifty, A., 2013. Proposed Seasonal Autoregressive Integrated Moving Average model for Forecasting Rainfall pattern in the

- Navrongo Municipality of Ghana, Journal of Environment and Earth Science 3(12), 80-85.
- Saurral, R., Barros, V., Lettenmaier, D.P., 2008. Land use impact on the Uruguay River discharge. Geophysical Research Letters. Vol. 35, 1-6.
- Stoffer, D., 2016. Applied Statistical Time Series Analysis. Package 'astsa'. <http://www.stat.pitt.edu/stoffer/tsa4/>.
- Torres, J.L., García, A., De Blasa, M., De Francisco, A., 2005. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy. Vol. 79, Issue 1, 65-77.
- Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology 239, 132-147.
- Tucci, C., 2009. Some scientific challenges in the development of South America's water resources. Hydrological Sciences Journal. Vol.46, 937-946.
- Wei, W., 2005. Time series analysis, univariate and multivariate methods, Second edition. Addison Wesley Longman Inc. Div Pearson, 1-624.
- Yamamoto, R., Iwashima, T., Sanga, N., Hoshia, M., 1987. Detection of the climatic jumps. XIX IUGG. General Assembly. August 9-22. Vancouver, Canadá.

Este es un artículo de acceso abierto distribuido bajo la licencia Creative Commons, que permite el uso ilimitado, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada.